DOCUMENT RESUME

ED 247 551 CS 207 488

AUTHOR Cronin, Mary C.
TITLE The Centrality of Testing to the Theory of Reading Processes.
PUB DATE [82]
NOTE 34p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Learning Theories; Reading Diagnosis; Reading Instruction; *Reading Processes; *Reading Tests; *Test Construction; *Testing; *Test Validity
IDENTIFIERS *Theory Development

ABSTRACT

Noting that in recent years testing has fallen into disrepute, this paper demonstrates how the construction, use, and validation of tests, if adequately conceived, can increase the level of knowledge about the intricate nature of the reading processes. The first section of the paper discusses the measurement process and the process of test validation, and examines reasons why tests have often failed as instruments of scientific inquiry in reading. In addition, the section describes the Assessment of Reading Processes (ARP), a construct-referenced measure that views reading as a constant exchange between reader and text. The second section of the paper presents a theory of assessment that defines (1) the limits of the reading processes, (2) how such covert mental operations may be inferred from the reading product, and (3) rules for the generation of the structure and content of assessment items. Other topics discussed in this section include the assessment situation, the reader-text interaction, and the cognitive processes involved in reading. The third section offers a model of construct-referenced test construction and discusses each of the components of such a model, while the last section discusses the construction of the ARP. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247551

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Mary C. Cronin
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE CENTRALITY OF TESTING TO THE THEORY OF READING PROCESSES

Purpose and Background

The purpose of this study is to show how the construction, use, and validation of tests, if adequately conceived, can increase our knowledge about the intricate nature of the processes involved in reading. This thesis is but a logical consequence of the circular and mutually dependent relation of measurement and theory in science. In spite of its sometimes negative connotations, "testing" is simply "the means by which educators measure" (Maguire, 1981, p. 10-1) ; for the remainder of this paper, "testing," "assessment," and "measurement" will be used interchangeably.

Scientific knowledge is a constant exchange between people and nature ; even in the so-called "hard" sciences, discoveries are perceived in the mind as much as in empirical terms (see Bronowski, 1973 ; Kuhn, 1962/1970). In the social sciences almost all measures are of "constructed" concepts, e.g., intelligence, achievement and reading comprehension. Hence, it is impossible to make any claims about a measurement without some underlying theoretical network. This is true even when measuring rigorously defined narrow sub-skills of reading ; because

CS207488

as Walmsley (1979), who created a criterion-referenced test of CVC pronunciation, points out :

> it is possible to rigorously describe an aspect of the reading process (to define a reading universe), but the definition will necessarily be derived from the test constructor's perspective on the reading process. Reading universes do not exist outside of this context (p. 602).

Thus, a reading theory, whether articulated or embedded in the test constructor's reality set, is inextricably entangled in the measurement process. Hence, theoretical claims must be made to validate a measure of reading, because it must be demonstrated that the instrument is measuring a particular concept, and further that this concept has some basis in reality. This conception of validity is called construct validity :

> Construct validation takes place when an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconforming the claim (Cronbach and Meehl, 1967, p. 65).

This process of test validation is, therefore, identical to the scientific process where a theory and the means of measuring it are simultaneously accepted, questioned, or rejected ; furthermore, this notion of validity rests on a well-attested position within the philosophy of science

(e.g., Freigl, 1953 ; Hempel, 1966 ; Kaplan, 1964). In recent years it has gained attention from psychometricians (e.g., Messick, 1975, 1980) and educators (e.g., Wardrop et al., 1978) alike : it is the only type of validity admissible if testing is to make fruitful contributions to reading theory (see Loevinger, 1967).

In the past, tests often failed as instruments of scientific inquiry in reading because two distinct groups of specialists -- reading theorists and psychometricians -- were engaged in developing tests. The wedding of knowledge from these two disciplines is one of the unique contributions of the present project in theory and test development, or construct validation.

First, a theory of reading was outlined ; then a tool to measure this conceptualization. This process is non-linear, and the theory was pre-shaped by the means available to assess that construct. At a later date, this measure will be used to test specific hypotheses derived from the theory ; in the meantime, however, it will help amass validational data of an exploratory nature. One such study -- a factor analysis -- has been carried out.

A test construction model guides this entire project. It embodies the systematized yet uncharted ways of scientific inquiry as it can incorporate future unforeseen investigations.

The present study and test of reading as processing will be successively refined until they are overtaken by a more apt concept of reading. Finally, this paper can offer but glimpses on the course of construct validation ; some events, e.g., test construction, will be detailed, others merely introduced.

The Measurement of Process

In the past two decades, many behavioral sciences, e.g., anthropology, biology, history, and psychology, have undergone a major paradigm shift (Kuhn, 1962/1970) from the study of structures and their taxonomies to the exploration of process or dynamic structures. Perhaps the appeal of process theories is their explanatory power or capacity for elucidating "relationships between otherwise distinct and independent clusters of phenomena" (Estes, 1979, p. 47). Thus, a process account is more efficient for understanding the nature of change and how it may be effected ; learning or effecting change is the business of educators.

In reading no less than in other fields, interest has shifted from the study of reading skills and achievement to understanding how the mind makes sense of print. Two strands within this trend are of immediate interest : since the mid-sixties a series of investigations into reading as a psycholinguistic process has been carried out "under the general direction of Kenneth S. Goodman" (Goodman and Burke, 1972, p.9). These studies explore the reading process (e.g., Biemiller, 1970 ; Weber, 1970) and refine its measurement

(e.g., Haupt and Goldsmith, 1982 ; Hood, 1976). Another grouping of theorists dubbed the discourse processing theorists (see Beebe, 1981) analyze the reading process by exploring the comprehension of connected discourse (e.g., Frederiksen, 1977 ; Kintsch and Van Dijk, 1978 ; Mandler and Johnson, 1977) ; unlike the psycholinguistic theorists, these latter researchers are not guided by a single research paradigm. However, both groups stress the role of the reader's background knowledge and the mental operations involved in matching it to information on the page. On the other hand, they differ on : (1) the unit of comprehension -- mainly sentence for the psycholinguists and passage for the discourse processing group ; and (2) the timing of the investigation. The present research synthesizes the work of these two groups of theorists to form a theory of reading processing.

It is apt, therefore, that recently construct validity has been seen as elucidating "the psychological processes required to complete the items on a test," and not "the kinds of information in the items" (Kirsch and Guthrie, 1980, p. 80) : the interaction between examinee and test determines performance, not the items per se. A process conceptualization overcomes many of the technical problems encountered in unequivocally demarcating a domain of content (see Loevinger, 1965) and referencing a test to that domain (cf. criterion-referenced tests). In contrast the present test, The Assessment of Reading Processes (ARP) will be "construct-referenced" (Messick, 1975, p. 95), i.e., constructed and interpreted

within the guidelines of a theoretical explanation of reading as a constant exchange between reader and text. A similar process method for the dynamic assessment of intelligence was developed by Feuerstein (1979) : intelligence is measured as a dynamic process rather than as an immutable substance ; hence the aim in testing is to gain insights into processing strengths and weaknesses, to guide instructional programs rather than to label the child. Likewise, the ARP is designed to paint a learner's processing profile to guide the teacher in program-planning, while making a valuable contribution to reading theory.

Theory of Assessment

The construct of reading explored in this work is broad and complex : it includes the entire process of gaining meaning from print. More specifically, this reader-text exchange is defined by the dynamic interplay of eight reading processes : attending, analyzing, sequencing, associating, predicting, monitoring, inferencing, and synthesizing. The theory of assessment of reading processes defines : (1) the limits of these processes, (2) how such covert mental operations may be inferred from reading product, and (3) rules for the generation of the structure and content of the assessment items.

Although the theory of reading processes may explain reading in a variety of circumstances, it is developed here within an assessment situation ; this contextualization will tell, both for theoretical and practical purposes, the

generalizability of such behavior.

Theory of Assessment of Reading Processes

Reading takes place whenever the reader interacts with the text, abstracting features (e.g., letters, words, or phrases), integrating these with his anticipations or world knowledge, allowing the text a certain autonomy while building a model of the message to fit his zeitgeist -- reading thus entails meaning and comprehension. The context of this encounter, however, affects the reader's emotional state, purpose, and hence processing.

Assessment Situation. The reading processes may be inferred from samples of oral and silent reading : as the child reads orally, they may be viewed through the miscue analysis technique (Goodman and Burke, 1972) ; however, a post facto inferral is possible from an analysis of unaided recalls following oral and silent reading. Thus, the measuring device is an informal reading inventory and entails an individualized assessment.

The salient components of the reading assesment situation are modeled in Figure 1 (following page) : reader-text interaction, i.e., processing, is influenced by the reader-examiner encounter and other situational variables ; this processing is inferred from oral reading and unaided recalls of oral and silent reading.

In an individualized assessment a good examiner-examinee rapport is essential : sex, race, age, and interactional style of the examiner may seriously affect the quality of the child's

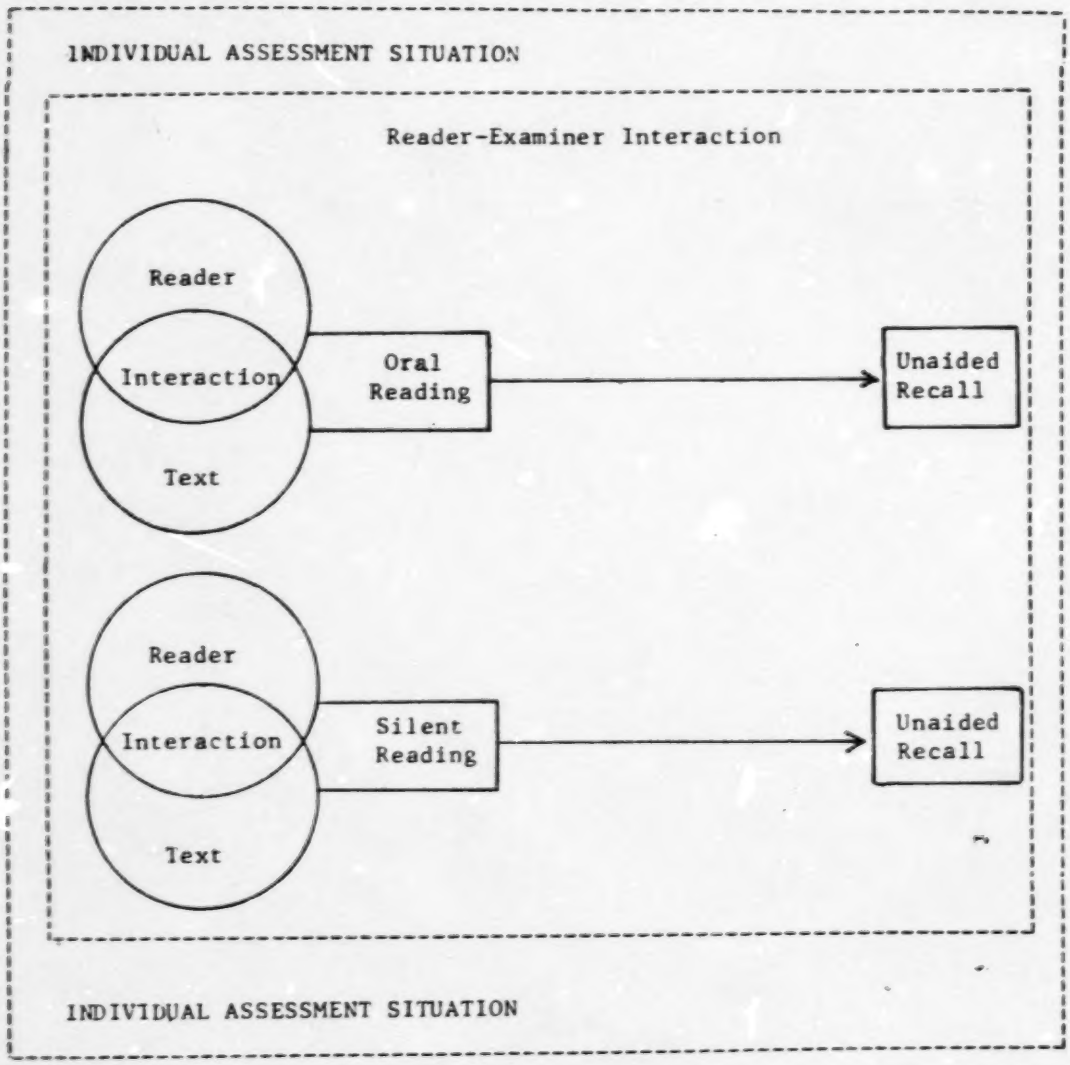

Figure 1. The Reading Assessment Situation

reading. A threatening environment or unfamiliarity with the task should also have negative effects. Such influences will be diminished by first involving the student in a practice assessment ; then if he/she accepts the situation as valid, the test directions should determine reading purpose.

Reader. In order to read, the reader must possess certain knowledge -- knowledge of the world, of language, and of the

reading task -- and relate it to the printed information. Since the focus of this analysis is on processing and not on assessing reader knowledge per se, prior knowledge should be controlled. Despite vast differences in background, most children who attend school in North America today have some experiences in common ; furthermore, research has shown that they possess a schema to comprehend conventional story structure (see Stein, 1978).

Text. Text, another basic component of the reading situation, may be analyzed into the following information-bearing components : graphics, vocabulary, syntax, text structure, and topic. Quality and quantity of processing will be affected by the interaction of these four aspects of text with the corresponding facets of reader knowledge. This should be considered when creating assessment items. For example, the subject matter should be as universal as possible ; most children are familiar with sports and common animals as a result of schooling and TV viewing.

Reader-Text Interaction : Reading Processes. The theories underlying the two methods of assessment -- miscue and recall analyses -- depict reading as a trade-off of information between reader and text,and conceptualize this exchange as cognitive and linguistic processing which occurs during and following reading. Miscue analysis (Goodman and Burke, 1972) may be extended and modified to enable processes such as attention to graphics or syntactic and semantic prediction to be inferred. Moreover, "retelling provides an overt measure

of what a student has done with the information presented in the selection" (Pearson and Johnson, 1978, p. 129), thus providing a window on such processes as inferencing and synthesis. Fagan's (in press) adaptation of the Drum and Lantaff (1977) recall categories is based on Kintsch and Van Dijk's (1978) work on the processes involved in text comprehension and recall ; each category of recalled information is either "reproduced," "transformed," or "reconstructed."

The eight processes inferred in this assessment are similar -- even identical -- to general cognitive processes ; however, each is defined and delimited in terms of its role in word identification and comprehension. The eight processing labels chosen are borrowed directly from on-going work at the University of Alberta, as exemplified in Malicky (1979). Nevertheless, these processing notions are present explicitly or implicitly in the work of many other reading theorists, e.g., Goodman (1976), Mandler and Johnson (1977), and Smith (1975, 1978).

1. Attending : This is the first contact a reader has with the graphics ; it refers only to the focusing of attention on graphic information.

2. Analyzing : Analysis refers only to the discrimination, organization, and abstraction of visual information from the instant the image hits the retina until recognition is made.

3. Sequencing : Sequencing is an adjunct process of attending-analyzing ; because print is linear, letters and words must be processed in an <u>overall</u> left-right direction.

4. Associating : Two types of association occur -- meaning and sound-symbol association. Sound-symbol association is involved when the reader uses a grapheme-phoneme correspondence to pronounce a word ; but meaning association is involved when the graphics "immediately" trigger off the meaning of a word or phrase (Smith, 1978). This latter association also functions in the storage and retrieval of larger units of text such as phrase or clause : units of text may be memorized in a verbatim manner and "reproduced" in recall, or the gist meaning may be abstracted and recalled with a "transformed" surface structure (Kintsch and Van Dijk, 1978).

5. Predicting : Although prediction may occur at many levels during reading, here it refers only to the use of semantic and syntactic knowledge to anticipate sentence meaning and word order.

6. Monitoring : In sustained reading, the processes of prediction and monitoring operate in a reciprocal relationship : "Having made predictions, the reader must use confirmation strategies to check the consistency of his expectations with cues he is encountering as he reads on" (Goodman, 1973, p. 62). Recently, monitoring has been studied more broadly as meta-cognition (e.g., Brown, 1980 ; Flavell, 1981). In this study, however, it will be inferred only from the correction of oral reading errors.

7. Inferencing : Inferences -- agents of integration and cohesion -- are made as the reader links new information from the text with previous information from the same text. How a

reader utilizes prior knowledge to "fill-in" a text, in part, explains why the same text may be understood in different ways by different readers. However, an inference is always substantiated by material stated by the author ; elaborations or embellishments of the author's schema are not inferences.

8. Synthesizing : Synthesis operates at three levels -- word, sentence, and passage. Word synthesis is the blending of phoneme, syllable, or morpheme units as a step towards word pronunciation ; sentence or phrase level synthesis is the chunking of information from these units for memory. Finally, passage-level synthesis involves the organization of the sub-parts of a discourse into a coherent structure. This process is the reader's main tool when committing text to long-term memory.

Interaction of the Processes : No process can operate in isolation ; when reading, the reader simultaneously processes graphic, syntactic, and semantic information (Goodman, 1976). Hence, the processes are inextricably intertwined, e.g., analysis of graphics is influenced by the extent to which the reader is able to predict the communication. Nevertheless, for the purposes of analyzing and measuring the reading process certain mental operations may be isolated.

Conclusion

Finally, the processes selected for measurement and the issues discussed are not an exhaustive account of reading ; other aspects of processing and the reading situation await future expansions of the theory and refinements of the ARP.

The theory of reading processes is but one side of the test-construction enterprise ; the other is a model of test construction.

A Model of Construct-Referenced Test Construction

Figure 2 below presents the model of test construction which guides this on-going work in theory and test validation.

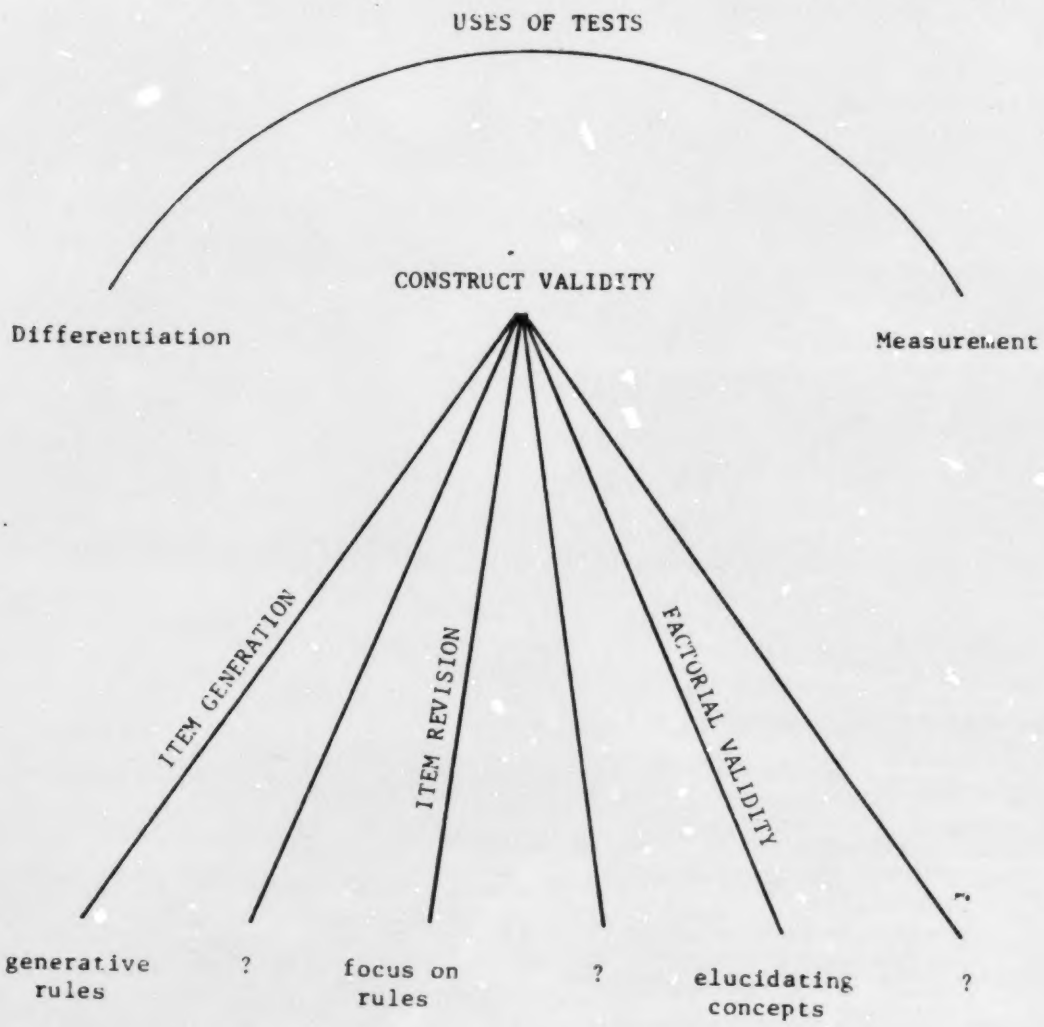

Figure 2. Model of the Construction of Construct-Referenced Tests

This is an adaptation of the Wardrop et al. (1978) model ; the revisions incorporate the theory of reading processes into all phases of instrument construction. In this way both instrument and the traits being measured are referenced to a theory-based construct of reading processes. It is thus uniquely suited to measuring how an examinee engages in a particular task (The psychometric rationale is developed in Cronin, 1982, pp. 79-115).

The first dimension, test uses, forms an umbrella under which all major decisions are made, viz., how will the assessment of reading processes be constructed and validated ? In the second dimension, construct validity is the overriding concern ; it subsumes all facets of validation (and reliability), as well as the procedures employed in constructing and revising assessment items. Because setting up construct validity is an unending process, this dimension consists of an infinite number of possible sub-ordinate nodes ; to date, three are named -- item generation, item revision, and factorial validity. Others will be identified in future research.

The test construction model represents other unique aspects of this study : (1) it is an explicit delineation of construct-referenced test construction, (2) which incorporates the theory of reading into every aspect of the construction of the ARP -- in other models (cf. Wardrop et al., 1978), validation only becomes an active concern after item construction ; and (3) construct validity is conceived as an unending process.

Construction of the ARP

The dimensions of the model of test construction guided

the building of the Assessment of Reading Processes (ARP) :

Test Uses

Since the ARP is an instrument capable of measuring the extent to which a reader engages in the eight reading processes it therefore falls on the right-hand side of the uses dimension ; hence it is an absolute measure of processing. And since these hypothetical constructs are not observable directly, a theoretically-based inference system (Wardrop et al., 1978) must be incorporated into the ARP.

Construct Validity

Two of the three identified stages of construct validation -- item generation and revision -- are of immediate concern :

Item Generation. Four distinct stages in the generation of domain-referenced items were named by Anderson et al. (1978). Here they are adapted for use with construct-referenced items :

1. Domain Identification : The domain involves the reader-text interaction when the reader is reasonably familiar with the structure, topic, syntax, and vocabulary of the text. More specifically, it includes the processes -- attending, analyzing, sequencing, associating, predicting, monitoring, inferencing, and synthesizing -- which occur when a Canadian elementary school child reads a narrative following a familiar structure on the topic of sports or animals.

2. Prototypic Items : An informal reading inventory is the means of assessment and consists of : a series of graded narrative passages (grades 1 to 6), directions for test administration, and a scoring system for the miscue and recall

analyses.

3. Variables Affecting the Difficulty of Items : The reading theory explicates four variables which should affect passage difficulty. These variables generate ten criteria or rules for passage writing :

(i) Passages will be written on sport and animal topics ; they will increase in difficulty by progressing from the more universal to the more specialized sports, and from the more common to the more exotic animals.

(ii) The number of words per passage will increase across the six grade levels.

(iii) Certain key content words will appear across each passage (for referential coherence, see Kintsch and Van Dijk, 1978).

(iv) The choice of lexical items will range from very familiar and concrete to less frequent and abstract words. Vocabulary will also be scaled in terms of the word frequency index in the American Heritage, Word Frequency Book (Carroll et al., 1971).

(v) Each passage will contain some "difficult words, i.e., unusual words or words with irregular spelling patterns (to facilitate miscue analysis).

(vi) As passages increase in difficulty, so will the average number of words per t-unit.

(vii) The number of semantic propositions per passage will increase across grade levels.

(viii) The Mandler and Johnson (1977) story grammar will be

used : passages will increase in difficulty by adding an extra character and episode every two grades.

(ix) The level of relative difficulty will be checked using the FOG readability index.

(x) A neutral title simply indicating topic will be given to each passage. This will provide the reader with a prior set or framework to comprehend each story. Unlike the other variables, the title is not intended to increase difficulty across levels.

4. Creating Sample Items : At this juncture, passages for two parallel forms of the ARP and scoring systems were created. The initial passages were written -- six on sports and six on animals -- according to the criteria above. Two scoring systems were divised : the Goodman and Burke (1972) miscue analysis was modified to permit the inferral of processes. This dimension was also added to the Fagan (in press) recall categories ; these categories plus the processes which may be inferred from each are indicated on Table 1 (following page).

Item Revision. In order to evaluate the criteria for writing the stories and the efficiency of the scoring systems for measuring a child's processing strategies, forms A and B of the ARP were piloted with twenty-four good and poor readers from grades one to six. It could also be seen whether both forms involved the reader in similar levels of processing.

Story Revision : The children's oral reading and recalls

Table 1
Recall Categories

TEXT EXACT	A
Exact recall of text • Associating--textual information "reproduced"	
TEXT SPECIFIC	B
Specific references in text • Associating, sequencing, synthesizing--textual information "transformed"	
TEXT ENTAILED	C
Summary statements • Synthesizing (across clauses)--information from within text "reconstructed"	
TEXT EXPERENTIAL	D
Experential addition to text, e.g., inferences • Inferencing--related information from outside the test "reconstructed"	
TEXT ERRONEOUS	E
Information which has been incorrectly processed • Faulty associating, faulty inferencing, faulty synthesizing	

were scored. A chart was then compiled for each story, noting the grade level of each child, followed by a list of his/her miscues plus comprehension problems. An excerpt from one of these charts for Track and Field (level 5) is given below :

Name	Grade-level	Miscues	Recall problems
Trevor	4; low reader	enthusiastic excells, routine, managed, although, competitions, imagining, unfortunately.	problem with separability of characters; thinks that Frank is practicing for the jumping and has the accident when competing.

These charts were examined to detect which stories parts were consistently too difficult or too easy for various

readers. Stories were also judged across and within levels : the former, to see if the stories presented a gradation of challenege to the reader ; the latter, to judge the parallelism of stories at each grade level.

Although nine of the passage-writing rules were generally satisfactory, some revisions were nonetheless necessary. They consisted of three categories :

1. Vocabulary : Some stories persistently caused more miscues then their partners at the same grade level ; sometimes these word identification problems were concentrated in one story part.

2. Structure : Since all children in the pilot study, regardless of reading ability, were familiar with the story structure, structural problems were minor. In some stories clearer separability of characters (Rubin, 1978) was required. Stories also needed to be equalized in terms of underlying and stated story parts, e.g., a stated moral was usually recalled, otherwise morals were omitted from recall ; this provoked inequalities in recall. Finally, referential coherence -- pronouns, synonyms, etc -- caused sone difficulties.

3. Conceptual Load : This refers to the ease with which children referenced stories to their past experience, and seems to bear on the number of diverse schemas a child had to employ to comprehend a story. For example, the level 3 stories had identical episodic structures and number of semantic propositons, yet one story happens in the home and the other tells of an accident at the swimming pool and involves a sick girl, a drowning boy, and a rescue. This aspect of semantic structure

was not specified in the rules, so an additional rule was specified : The range of different conceptual referents or schemata required to comprehend a story should be considered. At present, this is achieved intuitively since as yet there are no explicit guidelines.

Even though the three revisions were introduced separately, all intertwine and co-determine word identification and coherence problems ; no single aspect occurring in isolation caused major differences in levels of processing.

Only one of the rules for generating items was dropped : the FOG Readability Index. As the variables it controlled -- word length and sentence length -- did not predict item difficulty, it is noteworthy that this rule was not based on the theory of the assessment of reading processes.

Rewriting Stories and Writing Forms C and D. When the interaction of the problematic aspects of the stories was understood, forms A and B were revised, and forms C and D written -- A and C on sports and B and D on animals.

Revision of the Miscue Analysis. Revisions were made to the initial miscue analysis so as to (1) eliminate categories -- e.g., coding errors as high, average or low attention to graphics -- which lacked precision in measuring processes, (2) introduce new categories or processing indicators, and (3) establish an outside criterion of the child's comprehension.

The revised miscue analysis scoring sheet is presented in Figure 3 (following page). The columns are marked from 1 to 18. Fifteen of these are processing indicators, or in other words, processes are inferred from the information in these

Text	Miscue	Corrected + / -	Text word length in graphemes	Graphic cues accessed	Graphic cue position I M F	Cue sequence + / -	Text word length in phonemes	s/s association phonemes	s/s association position I M F	real/non word accessed RN	unit accessed Ph. W. Pa.	Oral language (reader s meaning)								Author's meaning		
												Syntax		Semantics			Prediction					
												(a) Sent.	(b) Part.	(a) Sent.	(b) Part.	(c) Pass.	(a) Sent.	(b) Part.		No change	Minor	Major
1	2	3	4	5	6	7	8	9	10	11	12	13		14			15			16	17	18
		Monitoring		Attending	Attending/analysing	Sequencing graphemes		Sound-symbol associating	Phonemic chunking/analysing/sound associating	Meaning associating	Analysing/meaning and sound association	Syntactic predicting	Syntactic predicting	Sentence meaning synthesizing	Partial sentence meaning synthesizing	Passage level synthesizing	Sentence syntactic and semantic predicting	Partial sentence syntactic and semantic predicting				

Figure 3 : Revised Miscue Analysis Sheet

columns. These a e indicated directly on the sheet.

Examples of how 11 miscues are scored appear in Figure 4 (following page), and where these occur in a child's reading of Track a.d Field is marked on Table 2 below :

Table 2

Example of coding oral reading errors.
The key words are underlined.

Track and Field

Frank, Mary and their friends are interested in track and field. Frank excells at running and Mary at the high jump. The friends are not real champions but practice hard to improve their sprinting (squating). Last Spring they decided they would enter all suitable events in the track and field competitions (com-pet-i-tions). Although they followed a strict training routine (routaine) for a month leading up to the competitions, their plans did not turn out as expected. Mary and the friends practiced their jumping every day and did exercises (their) to strengthen their leg muscles. Frank was determined to become the junior champion of the one-hundred metre dash so it was important for him to move instantly on the sound of the starting gun. Every day he practiced his start by imagining himself pushed forward by a strong wind. At last the big day arrived. Mary was the first to compete. She expended all her energy but still only managed to get third prize for jumping. Their friends did not win even a single prize. Now all their hopes were on Frank to become the champion at the one-hundred metre dash. But after the jumping events, Frank thought he would try the high ju(ing) (the) for (of it) fun. When the area around the jump was deserted, Frank sprinted, sprang high in the air and cleared the bar. Unfortunately for him, he landed in the sand crying with pain. His ankle (angle) was badly (hardly) sprained. Mary and the friends came running. They helped him to the first aid room just as the one-hundred metre event was being (beginning C) announced.

A complete explanation of the scoring is given in Cronin (1982, pp. 140-154). For our purpose here, an example of coding

Text	Miscue	Corrected + / -	Text word length in grapheme	Graphic cues accessed	Graphic cue position I M F	Cue sequence + / -	Text word length in phoneme	s/s association phonemes	s/s association position I M F	real/non word accessed RN	unit accessed Ph. W. Pa.	Oral language (reader's meaning)							Author's meaning		
												Syntax		Semantics			Prediction				
												(a) Sent.	(b) Part.	(a) Sent.	(b) Part.	(c) Pass.	(a) Sent.	(b) Part.	No change	Minor	Major
1	2	3	4	5	6	7	8	9	10	11	12	13		14			15		16	17	18
(1) the	--	--										+	+	+	+	+	+	+	+		
(2) sprinting	spurting	--	9	7	IMF	--	8	6	F	R	W	+	+	+	+	+	+	+			
(3) competitions	compet-i-tions	--								R	Pa.	+	+	+	+	+	+	+	+		
(4) --	their	--								R	Ph.	+	+	+	+	+	+	+	+		
(5) win even	even win	--	7	7	IMF	--	7	7	IMF	R	Ph.	+	+	+	+	+	+	+	+		
(6) jump just for fun	jumping for the fun of it	--	4	4	IMF	+	4	4	IMF	R	Ph.	+	+	+	+	+	+	+	+		
(7) routine	routaine	--	7	7	IMF	+	5	4	IF	N	W	+	+	-	-	+	-	-		+	
(8) sprinted	spr--in	--	8	5	IM	+	8	5	I	N	Pa.	-	-	-	-	+	-	-		+	
(9) ankle	angle	--	5	4	IMF	+	4	3	IF	R	W	+	+	+	+	-	+	+			+
(10) badly	hardly	--	5	4	MF	+	5	4	F	R	W	+	+	+	+	-	+	+		+	
(11) being	beginning	+																			
TOTAL	11	10																			

Figure 4 : Examples of Errors Coded

and inferral of processing plus an explanation of the data in columns 16 to 18 will be given.

Column 5 : This column contains a record of the graphemes in a word the reader attended to. The miscue is compared to the stimulus in the text to discover how many letters the reader attended to to utter the miscue. Any possible pronunciation of a grapheme is accepted as a sign of attending, e.g., when "slough" is pronounced "slow" the child is given credit for attention to all graphemes, as /o/ is an acceptable association for the "ough" spelling pattern. A score on attending is calculated as a proportion of the total letters in the text word. These scores may be used to depict a reader's processing and to make such generalizations as : When reading narrative material at a grade-four level, Ann attends to print 90 per cent of the time while associating meaning to less than half of these miscues. This information can help direct instruction, but an overall proportion-of-processing score is used in quantitative research.

Columns 16 to 18 : The data collected in these columns act as an outside criterion for the reader's processing and comprehension of the text. Columns 16, 17, and 18 respectively indicate whether the miscue has caused no change, minor change, or major change in the author's intended meaning. No Change in meaning occurs when the message intent is not altered in any significant way, e.g., changes in function words : "he saw horses in the field" for "he saw the horses in the field." However, context must be considered : an insignificant change in one context can be highly significant in another. Minor

meaning changes alter only the meaning of the immediate sentence and not the gist of the passage. However, this is also determined by a more rigid criterion -- a minor change cannot affect the intended meaning of the key content words (1/3 of the text); these are underlined in the story on Table 2 (p. 22 above). Finally, an error is scored as a major meaning change when it significantly alters the meaning of a key word.

Revision of the Recall Categories. Only two minor changes were made to the recall analysis : (1) information in the Text Experiential (D) category was divided so as to distinguish between textually-constrained inferences (D_1) and recalled elaborations triggered off by the theme of the story (D_2) ; and (2) the Text Erroneous (E) category was split to distinguish between two types of faulty processing. An analysis of a recall protocol, including the revisions noted above, is given on Table 3 (following page).

Summary

The theory of reading assessment guided the construction, piloting, and revision of the ARP -- a measure of the reader's processing. For the moment, reader is a Canadian elementary school child, text is a story on the topic of sport or animals, and the assessment situation includes a training and testing component, and the researcher as the examiner.

The next phase in validation was a factor analysis. For reasons of space it is not possible to report this study in its entirety -- for further details see Cronin (1982, pp. 158ff.). Since this was the first practical application of the revised

Table 3

Sample protocol indicating coding into recall categories

Track and Field

Frank and Mary are interested[A] in track and field// and they to do the competitions and[C] be champions in the next track and field/ join all[B] events// they were practicing[B] every day// their friends were practicing[B] all the time doing special exercises // Mary was practicing[D] to do her high jump// and then the[B] big day came// and she won[B] third prize// by the end of the day they[E2] all won prizes// then Frank wanted to go high jumping over that bar[C] just for the fun of it// after he[E1] wasn't hurt// but they took him to[B] the first aid room// (then the nurse no first and then) and then the[B] hundred metre dash was announced// so he couldn't run[D1] in it// so the nurse gave[D2] him a check-up instead.

ARP, further refinements to the coding systems and passages in the near future seem inevitable.

Conclusion

In recent years testing, especially of the norm-referenced variety, has fallen into disrepute. The present work is an essay to reverse this drift by demonstrating that a holistic project in the measurement of reading processes can at once chart the course of reading instruction and scientific inquiry. Nevertheless, this mandate is only possible if construct validity is an overriding concern : measures must be conceived, created and revised in terms of a theory, but in turn they must be modified to fit empirical evidence collected by the assessment. Only the

first half of this procedure has been presented ; a factoral analytic study was carried out where 102 children read six passages -- a step in the opposite direction.

The theory depicted reading as interaction involving the constant trade-off of top-down and bottom-up processes (cf. Rumelhart, 1977). The results of the factorial study, however, suggest that while processes are interactive, this interaction may be structured by the unit of print processed. For instance, grapho-phonic processing -- attending, analyzing, and associating -- grouped on one factor, while processing inferred through recall tended to cluster on distinct factors. Though not unequivocal, these findings indicate some conceptual modifications -- perhaps a model of processing structured by the textual unit in immediate focus. Yet, further research is in order before a definite conclusion is reached : these results may be the consequences of the stepwise fashion in which processing was measured.

Heretofore, this project has shown the construct of reading as eight processes to have some basis in reality. But more pertinently, it is an exemplar or inspirational study to encourage theory-based measurement by showing that : (1) measurement of reading is logically impossible without a theory ; (2) complex explanations of reading can be incorporated into test designs, e.g., rules for passage writing ; and (3) a project in theory validation can operate at two levels simultaneously -- portraying an individual child's reading while building theory. Thus the centrality of theory to testing

and of testing to theory was advocated by an example -- the construction of the ARP -- rather than by rhetoric.

References

Anderson, T.H., Wardrop, J.L., Hively, W., Muller, K.E., Anderson, R.I., Hastings, N.C., and Fredericksen, J. Development and trial of a model for developing domain referenced tests of reading comprehension. Technical Report No. 86. Urbana, Illinois, 1978.

Beebe, M.J. A model of the relationships between reading strategies and reading comprehension. Unpublished doctoral dissertation, University of Alberta, 1981.

Biemiller, A. The development of the use of graphic and contextual information as children learn to read. Reading Research Quarterly, 1970, 6, 75-96.

Brownowski, J. The ascent of man. London: British Broadcasting, 1973.

Brown, A.L. Metacognitive development and reading. In R.J. Spiro, B.C. Bruce and W.F. Brewer (Eds.), Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artifical intelligence and education. Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Carroll, J.B., Davies, P. and Richman, B. Word frequency book. New York: American Heritage Publishing Co., Inc., 1971.

Cronbach, L.J. and Meehl, P.E. (1955) Construct validity in psychological tests. In D.J. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Cronin, M.C. Construct validity and the measurement of reading processes. Unpublished doctoral dissertation, University of Alberta, 1982.

Drum, P.A. and Lantaff, R.E. Scoring categories for protocols. Paper presented at the Second Annual Language Conference, Boston, 1977.

Estes, W.K. On the descriptive and explanatory functions of theories of memory. In L. Nilsson (Ed.), Perspective on memory research. Hillsdale, N.J.: Erlbaum, 1979.

Fagan, W.T. Comprehension categories for protocol analysis. In W.T. Fagan, C. Cooper and J. Jensen (Eds.), Measures for Research and Evaluation in the English Language Arts, Volume 2. Urbana, Illinois: The National Council of Teachers of English, (in press).

Feuerstein, R. The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. Baltimore: University Park Press, 1979.

Flavell, J.H. Cognitive monitoring. In W.P. Dickson (Ed.), Children's oral communication skills. New York: Academic Press, 1981.

Frederiksen, C.H. Semantic processing units in understanding text. In R.O. Freedle (Ed.), Discourse production and comprehension, Volume 1. Norwood, N.J.: Ablex Publishing, 1977.

Feigl, J. Confirmability and confirmation. In P.P Wiener (Ed.), Readings in philosophy of science. New York: Scribner's, 1953.

Goodman, K.S. Strategies for increasing comprehension in reading. In H.M. Robinson (Ed.). Improving reading in the intermediate years. Chicago: Scott, Foresman, 1973.

Goodman, K.S. Reading a psycholinguistic guessing game. In H.S. Singer and R.B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1976.

Goodman, Y.M and Burke, C.L. Reading miscue inventory. New York: Macmillan, 1972.

Haupt, E.J. and Goldsmith, J.S. Expanding the factorial structure of reading errors: frequency, severity, and higher-order components. Paper presented at the American Educational Research Meeting. New York, 1982.

Hempel, C.G. Philosophy of natural science. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Hood, L. Qualitative analysis of oral reading errors: The interjudge reliability of scores. Reading Research Quarterly, 1976, 11, 557-598.

Kaplan, R.I. The conduct of inquiry. New York: Basic Books, 1964.

Kintsch, W. and Van Dijk, T.A. Towards a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Kirsch, I.S. and Guthrie, J.T. Construct validity of functional reading tests. Journal of Educational Measurement, 1980, 17, 81-93.

Kuhn, T.S. (1962) The structure of scientific revolutions (2nd edition). Chicago: University of Chicago Press, 1970.

Loevinger, J. Person and population as psychometric concepts. Psychological Review, 1965, 72, 143-155.

Loevinger, J. (1957) Objective tests as instruments of psychological theory. In D.J. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Mandler, J.M. and Johnson, N.S. Rememberance of things parsed: story structure and recall. Cognitive Psychology, 1977, 9, 111-151.

Maguire, T.O. Measurement and evaluation. In J.H.M Andrews and W.T Rogers (Eds.), Canadian research in education: a state of the art review. Report prepared for the Social Sciences and Humanities Research Council of Canada, 1981.

Malicky, G.M. A reading processes approach to diagnosis and remediation. Unpublished paper, University of Alberta, 1979.

Pearson, P.D. and Johnson, D.D. Teaching reading comprehension. Toronto: Holt, Rinehart and Winston, 1978.

Messick, S.A. The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 1975, 30, 205-16.

Messick, S. Test validity and the ethics of assessment. American Psychologist, 1980, 35, 1, 012-1, 027.

Rubin, A.D. A Theoretical Taxonomy of the Differences Between Oral and Written Language. Technical Report No. 35, Center for the Study of Reading. Urbana, Illinois: University of Illinois, 1978.

Rumelhart, D.E. Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance, VI. Hillsdale, N.J.: Lawrence Earlbaum, 1977.

Smith, F. The role of prediction in reading. Elementary English, 1975, 52, 305-311.

Smith, F. Understanding reading. Toronto: Holt, Rinehart and Winston, 1978.

Stein, N.T. How children understand stories: A developmental analysis. Centre for the Study of Reading. Technical Report No. 69. Urbana, Illinois: University of Illinois, 1978.

Walmsley, S.A. The criterion-referenced measurement of an early reading behavior. Reading Research Quarterly, 1978, 14, 574-604.

Wardrop, J.L., Anderson, T.H., Hively, W., Anderson, R.I. Hastings, N.C. and Muller, R.E. A framework for analysing reading test characteristics. Technical Report No. 109. Urbana, Illinois: University of Illinois, 1978.

Weber, R. A linguistic analysis of first-grade reading errors. Reading Research Quarterly, 1970, 5, 427-451.